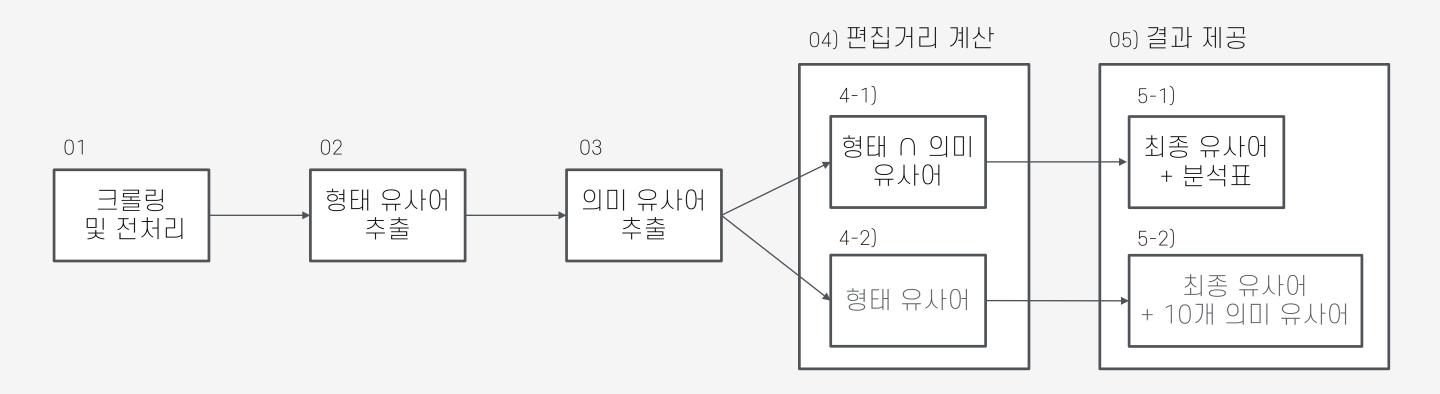
# 온라인 커뮤니티 특화 감성 사전 구축을 위한 새로운 용어 극성값 분석 시스템

백데이터미네이터

## 진행 사항

- 1. 학회 발표 동영상 및 드림패스 동영상 제출
- 2. 특허 아이디어 구체화 및 TEST

- KEY POINT
- ⇒ 온라인 형태 변형어의 편집거리 기반 형태 및 의미 유사어 분석 시스템



- 1) "네이버 블로그 "에서 사용자가 입력한 형태 변형어를 검색하여 데이터 크롤링 및 전처리
- 2) 입력한 형태 변형어의 형태 유사어 추출 (문장 단위 & 자소 단위 학습)
- 3) 입력한 형태 변형어의 의미 유사어 추출
- 4) 추출된 유사어들의 편집거리 계산
- 5) 최종 유사어 및 편집거리 분석표 제공

+

1) "네이버 블로그 "에서 사용자가 입력한 형태 변형어를 검색하여 데이터 크롤링 및 전처리

(예시) 형태 변형어 '댕댕이' 검색



#### 문장 단위 학습

- 1 평택강아지분양 귀여운 댕댕이 둘째 단모치와와
- 2 요 근래 업무량이 늘어남과 동시에 점점 출장이 잦아지면서 집에 있는 반려견이 계속 걱정이 되었습니다. 물론 일을
- 3 안산강마지부양 땡땡이와 힐링을 막다죠
- 4 아담스펫강아지&가필드고양이인천광역사 남동구 논현로번길 이번에 제가 안산강아지분양 통해서 더 나아진 삶으.
- 5 귀여운 댕댕이, 시츄 성격과 털빠짐 주의점
- 6 사자 라는 이름의 유래를 가지고 있는 시츄는 이름에서도 알 수 있듯 중국에서 온 강아지랍니다 체고는 목무게:
- 7 동탄강아지부양 매력넘치는 맨덴이
- 8 미니펫 강아지고양이 강남점서울특별시 강남구 논현로길 부모님의 반대가 심하셔서 강아지를 계속 못키우다가 언니?
- 9 종근당바이오 라비벳 장&구강 유산균 댕댕이 위해 골랐죠
- 10 우래기들이 어릴 땐 하루 종일 정신없이 움직여서 다들 에너자이저인 줄 알았어요어디든 저보다 훨씬 먼저 뛰어 올리
- 11 천안강마지분양 핸복을 주는 맹맹미들!
- 12 예전부터 강아지를 정말 좇아했는데출근하면 혼자 있어야하는 시간도 길고집도 넓지않아 여건이 되질 않았어요.그런!
- 13 프렌치불독분양 귀엽고 사랑스러운 막내댕댕이 데려왔어요
- 14 최근 부모님도 퇴직을 하시고 '저와 언니도재택근무를 시작하면서 가족들이 집에서보내는 시간이 많이 늘어났어요.
- 15 [국개대표1 기호성 좋은 사료로 맹맹이들 입맛 사로잡기
- 16 코로나 때문에 잘 먹던 웰니스 코어 라인이 수입이 좀 원활하지가 않더라구요. 미국사료협회 가이드라인까지 찾아
- 17 강아지 혈통 분석, 우리 댕댕이에게 어떤 피가 흐를까
- 18 시스템으로 강아지 혈통 분석 해주는 앱이 있길래 재미삼아 해봤다.우리 꼬미는 과연 어떤 결과가 나올까 혈통분

#### 자소 단위 학습

## 2) 입력한 형태 변형어의 형태 유사어 추출

#### 2-1) 문장 단위 학습

```
#문장 학습 - 형태 유사이 추출
model1 = fasttext.load_model('data_result.bin')
model1.get_nearest_neighbors("댐댐이")

Warning: `load_model` does not return WordVecto

[(0.9838256239891052, '사실'),
  (0.9823348522186279, '강아지를'),
  (0.9820657968521118, '댐댐이는'),
  (0.9797086715698242, '댐댐이들'),
  (0.9775006175041199, '경우'),
  (0.9774555563926697, '강아지와'),
  (0.9769712686538696, '댐댐이가'),
  (0.9764970669746399, '강아지는'),
  (0.9742645025253296, '강아지'),
  (0.9709970355033875, '있는')]
```

#### 2-2) 자소 단위 학습

```
from soynlp.hangle import compose, decompose, character_is_korean
 3 #자소단위 학습 - 형태 유사어 추출
 4 | model3 = fasttext.load_model('re_result.bin')
 5 model3.get_nearest_neighbors("앰앰이")
Warning: `load_model` does not return WordVectorModel or SupervisedModel any more,
[(0.32594582438468933, '¬ ⊦•≖ +|•• +|•• +|•• +|•• |,
 (0.25433990359306335, 'Laretell'),
 (0.2458125650882721. \Box \pm o ...).
 (0.1965695321559906, '∈ + ≥ .*'),
 (0.19008754193782806, '□ | ★*'),
 (0.16549181938171387, '¬¬¬¬¬¬¬¬¬¬¬¬¬,
 (0.1537555456161499, 'ㅂ ト•ㄹㅗ•'),
 (0.13369370996952057, '÷⊥∟•⊼ +•'),
 (0.12780514359474182, '¬非太工¬•'),
 (0.10005773603916168, '# | L * = 1 * ¬ 1 L * ')]
[hgtk.text.compose(found) for found in [word for score, word in model3.get_nearest.
['카페에', '놀러', '명', '털', '및', '조금', '바로', '혼자', '계속', '반려견']
```

#### 3) 입력한 형태 변형어의 의미 유사어 추출

```
print(loaded_model.wv.most_similar("댕댕이", topn=10))
[('댕댕이삼계탕', 0.9992862343788147), ('강아지삼계탕', 0.9992821216583252), ('강아지옷', 0.9992791414260864), ('퓨어메라', 0.9992114901542664), ('이벤트', 0.9991117119789124), ('강아지간식', 0.999061107635498), ('강아지', 0.9990420937538147), ('강아지옷만들기', 0.9990110397338867), ('중입니다.분', 0.9989457130432129), ('애견용품', 0.998928427696228)]
```

4) 추출된 유사어들의 편집거리 계산

## 4-1) 의미 유사어가 형태 유사어 list 단어에 포함되는 경우, 해당 단어들만 편집거리 계산

→ 형태 변형어 원형에 가장 가까운 단어를 추출하기 위해, 두 list를 crosscheck 하여 겹치는 단어 분석

#### 분석 단계

Idea 1) Str 길이가 같은 부분은 음소 단위로 편집거리를 계산한다.

Idea 2) Str 길이가 다른 삽입or삭제 음절의 경우, 해당 자리를 음소 단위로 편집거리를 계산하여 더한다.

Idea 3) 의미 유사어와 중복되는 음절의 개수만큼 편집거리에서 뺀다.

### 예人()

형태 유사어 list = {사실, 경우, <mark>강아지는, 강아지</mark>, 카페에, 놀러, 털, 혼자, 계속, 반려견}

의미 유사어 list = {강아지옷, 이벤트, 퓨어메라, 강아지간식, <mark>강아지</mark>, 강아지옷만들기, 애견용품}

## +

## 진행 사항 - 특허 아이디어 빌딩

### 4-1) 의미 유사어가 형태 유사어 list 단어에 포함되는 경우, 해당 단어들만 편집거리 계산

Idea 1) Str 길이가 같은 부분은 음소 단위로 편집거리를 계산한다.

⇒ 댕댕이 vs. 강아지는 → 2 + 2 + 1

Idea 2) Str 길이가 다른 삽입or삭제 음절의 경우, 해당 자리를 음소 단위로 편집거리를 계산하여 더한다.

⇒ 는 "∟/\_/∟" 음소 단위 편집거리 3 더하기 → 2 + 2 + 1 + 3 = 8

Idea 3) 의미 유사어와 중복되는 음절의 개수만큼 편집거리에서 뺀다.

⇒ 강아지 음절 개수 3 빼기 → 8 - 3 = 5

<u>Aa</u> 형태 변형어	≣ 의미 유사어 중복	≣ 형태 유사어	를 편집거리 계산	■ 중복 음절 개수 빼기	를 최종 편집거리
댕댕이	중복 x	사실	3 + 3 + 2 = 8		
댕댕이	중복 x	경우	2 + 3 + 2 = 7 Idea 1 Idea 2	Idea 3	
댕댕이	강아지 중복	강아지는	2 + 2 + 1 + 3 = 8	8 - 3	5
댕댕이	강아지 중복	강아지	2 + 2 + 1 = 5	5 - <mark>3</mark>	2
댕댕이	중복 x	카페에	3 + 3 + 2 = 8		
댕댕이	중복 x	놀러	3 + 3 + 2 = 8		

- 5) 최종 유사어 및 편집거리 분석표 제공
  - → 최종 편집거리가 가장 낮은
  - " 강아지 " 를 최종 유사어로 제공

<u>Aa</u> 형태 변형어	■ 의미 유사어 중복	臺 형태 유사어	➡ 편집거리 계산	■ 중복 음절 개수 빼기	■ 최종 편집거리
댕댕이	중복 x	사실	3 + 3 + 2 = 8		
댕댕이	중복 x	경우	2 + 3 + 2 = 7		
댕댕이	강아지 중복	강아지는	2 + 2 + 1 + 3 = 8	8 - 3	5
댕댕이	강아지 중복	강아지	2 + 2 + 1 = 5	5 - 3	2
댕댕이	중복 x	카페에	3 + 3 + 2 = 8		
댕댕이	중복 x	놀러	3 + 3 + 2 = 8		

- 4) 추출된 유사어들의 편집거리 계산
- 4-2) 의미 유사어가 형태 유사어 list 단어에 포함되지 않는 경우
- 1) 형태 유사어 list에 편집거리 적용하여, 편집거리가 가장 낮은 유사어 제공 ldea 1) Str 길이가 같은 부분은 음소 단위로 편집거리를 계산한다. ldea 2) Str 길이가 다른 삽입or삭제 음절의 경우, 해당 자리를 음소 단위로 편집거리를 계산하여 더한다.
- 2) 최종 결과에서 상위 10개 의미 유사어 함께 제공

Aa 형태 변형어	돌 형태 유사어	☲ 편집거리 계산	표 최종 편집거리
댕댕이	사실	3 + 3 + 2	8
댕댕이	경우	2 + 3 + 2	7
댕댕이	강아지는	2 + 2 + 1 + 3	8
댕댕이	강아지	2 + 2 + 1	5
댕댕이	카페에	3 + 3 + 2	8

[('댕댕이삼계탕', '강아지삼계탕', '강아지옷', '퓨어메라', '이벤트', '강아지간식', '강아지', 강아지옷 만들기', '애견용품']

## 차주 계획

- 1. 특허 아이디어 TEST 정확도 향상을 위한 전처리 개선
- 2. 특허 명세서 초안 작성

# 감사합니다:)

**빅데이터미네이터** 

10

손윤식 교수님

주식회사 닥터송 송환구 멘토님